Semantic Priors for Intrinsic Image Decomposition (Supplementary)

Saurabh Saini
saurabh.saini@research.iiit.ac.in
P. J. Narayanan
pjn@iiit.ac.in

CVIT, KCIS,
International Institute of Information Technology-Hyderabad,
India

1 Semantic Features Analysis

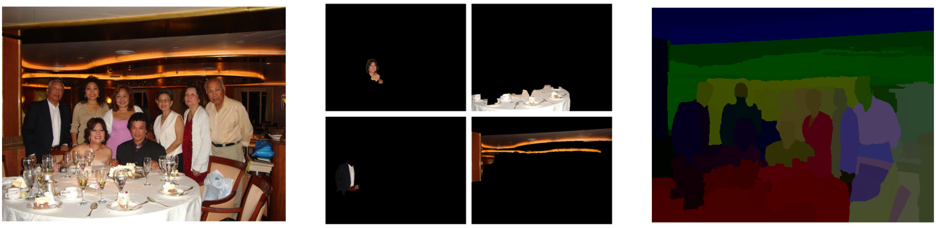

Figure 1: **Visualizing selective search features.** From left-to-right: Original image; four sample mask images from MCG (binary masks overlaid on the image for visualization) and dimensionality reduced image of selective search features (g_i) used for encoding class agnostic semantic information.

1.1 Visualization for Selective Search Features

Our selective search features are formed by concatenating various mask values at a particular pixel, weighted by MCG [1] 'objectness' score. We do dimensionality reduction on these features using PCA for efficient computation during reflectance formulation. We use dimensionality reduced features in *Stage 2* of the framework unlike *Stage 1* as the mid-level priors are re recomputed only in this stage in each iteration. We found no significant change in performance due to this during experimentation. Figure 1 shows a few sample sample masks (overlaid over the image for visualization) and the 'PCA-image' (formed by reducing the dimensions to 3) for an example image. Note how in the 'PCA-image' the regions belonging to the same object get clustered together illustrating how our selective search features, g_iś encode mid-level semantics.

1.2 Ablation Analysis

As mentioned in the main paper, in order to study the effects of various priors on the quality of Intrinsic Image Decomposition (IID), we experimented with several variants of our framework in an ablation analysis. The IID results obtained using these variants are shown

Figure 2: **v3 and v7 comparison.** From left to right: Original image, v3 reflectance and v7 reflectance respectively. Note how the color tone of reflectance results from v3 is incorrect although overall v3 shows better performance than v7 when evaluated using WHDR metric.

in Figure 4. Note, v1 has very little structural information as most of the shading priors are missing and hence derives results mainly based on color information. This causes incorrect IID reflectance as shown in column 1. v2 brings structural information in the form of S_g but in a few cases is unstable as no mid-level semantic information is present. v3 gives significantly better results compared to previous two as it has nearly all the priors but for several images leads to incorrect global reflectance tone (See image 3 and Figure 2) due to lack of global shading information. v4 and v5 give good reflectance results but do not handle shadows and lights well (See table shadow in image 2 and lamp in image 3). These are better handled by v6 due to our semantic prior R_m. As mentioned above, v7 has global context prior in addition to v3. This term although seems to degrade the framework performance by a small margin during the ablation analysis (v3 $WHDR = 18.15$ v.s. v7 $WHDR = 18.19$) but gives correct reflectance results as shown in Figure 2. This example also highlights the shortcoming of relative metric like WHDR.

We use the values of the parameters for Split-Bregman iterations as provided by Bi et al. [3] and empirically estimate the remaining parameters over a small subset of images. All analysis and results in our main paper and supplementary are generated using these fixed set of parameter values:

$\lambda_g = \lambda_m = 0.002$, $\lambda_l = 2$, $\gamma_g = 1$, $\gamma_m = \gamma_l = 20$, $\theta = 40$, $\tau = 1.2$, $t_c = 0.0001$, $t_m = t_b = t = 0.05$, $k = 5$

Figure 3: Sample images from IIW dataset.

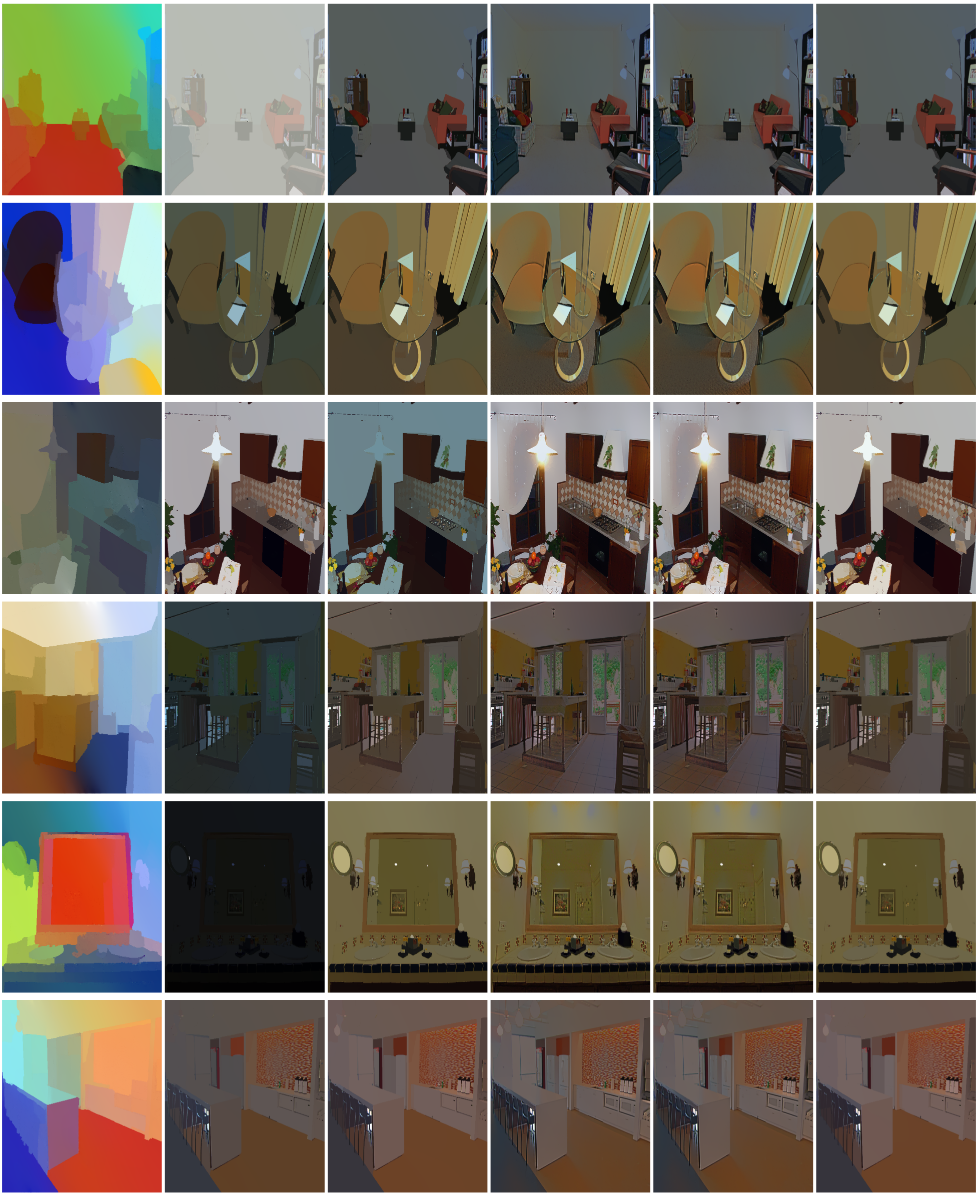

Figure 4: **IID using different variants.** In each scene from left-to-right: Results using variant v1, v2, v3, v4, v5 and v6 on images shown in Figure 3 (v7 is visually very similar to v6 and hence not shown). v3 gives good results as it contains all except one prior but does not handle global lighting well and in some cases leads to incorrect reflectance colour tone (See lamps in 3^{rd} and 4^{th} images). v4 and v5 lack mid-level reflectance sparsity term and is unable to remove the highlights from the scene (See how in 4^{th} and 5^{th} columns light gradients and shadows are not removed). v6 introduces our semantic mid-level reflectance sparsity and leads to better reflectance over v4 and v5.

2 Additional IIW Results

In this section we provide additional results on IIW dataset introduced by [2]. A few sample images from this dataset are shown in Figure 3. Below all results are arranged from left to right as: Original image, reflectance and shading respectively in Figure 5, Figure 6, Figure 7, Figure 8 and Figure 9.

Figure 5: **Results on IIW dataset.** In each row: Original image, reflectance and shading respectively.

Figure 6: **Results on IIW dataset.** In each row: Original image, reflectance and shading respectively.

Figure 7: **Results on IIW dataset.** In each row: Original image, reflectance and shading respectively.

Figure 8: **Results on IIW dataset.** In each row: Original image, reflectance and shading respectively.

Figure 9: **Results on IIW dataset.** In each row: Original image, reflectance and shading respectively.

3 Qualitative Comparisons

We show additional qualitative comparisons with Bell et al. [2], Zhou et al. [4] and Bi et al. [3] respectively, from left-to-right in Figure 10, Figure 11, Figure 12, Figure 13 and Figure 14. Our results are shown in the last column. Notice the persistence of noise and artifacts in the results of the other three methods. Also our method is able to handle light sources and soft shadows in the images better than the other methods.

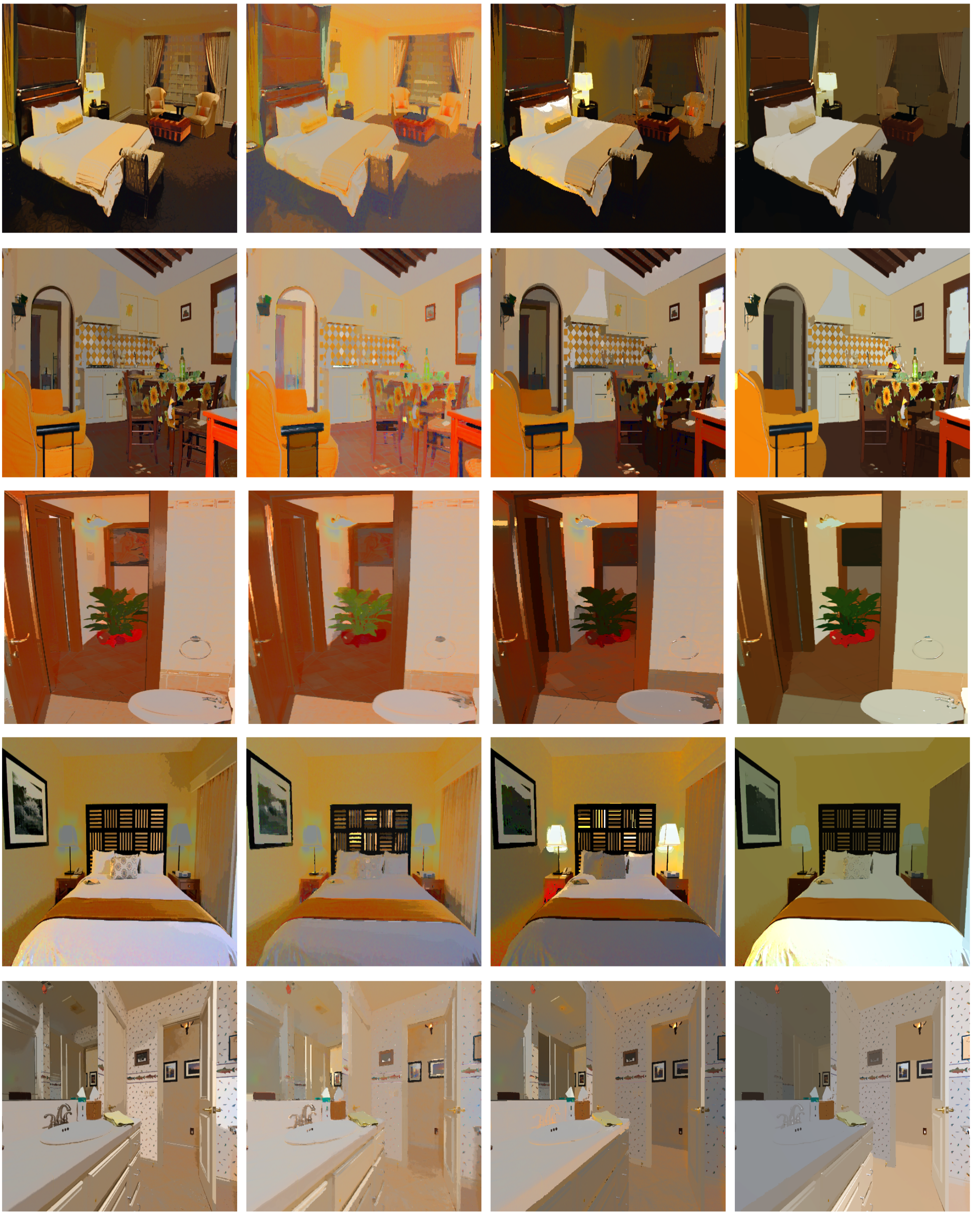

Figure 10: **Qualitative Comparison:** In each row results by [2], [4], [3] and our method respectively.

Figure 11: **Qualitative Comparison:** In each row results by [2], [4], [3] and our method respectively.

Figure 12: **Qualitative Comparison:** In each row results by [2], [4], [3] and our method respectively.

Figure 13: **Qualitative Comparison:** In each row results by [2], [4], [3] and our method respectively.

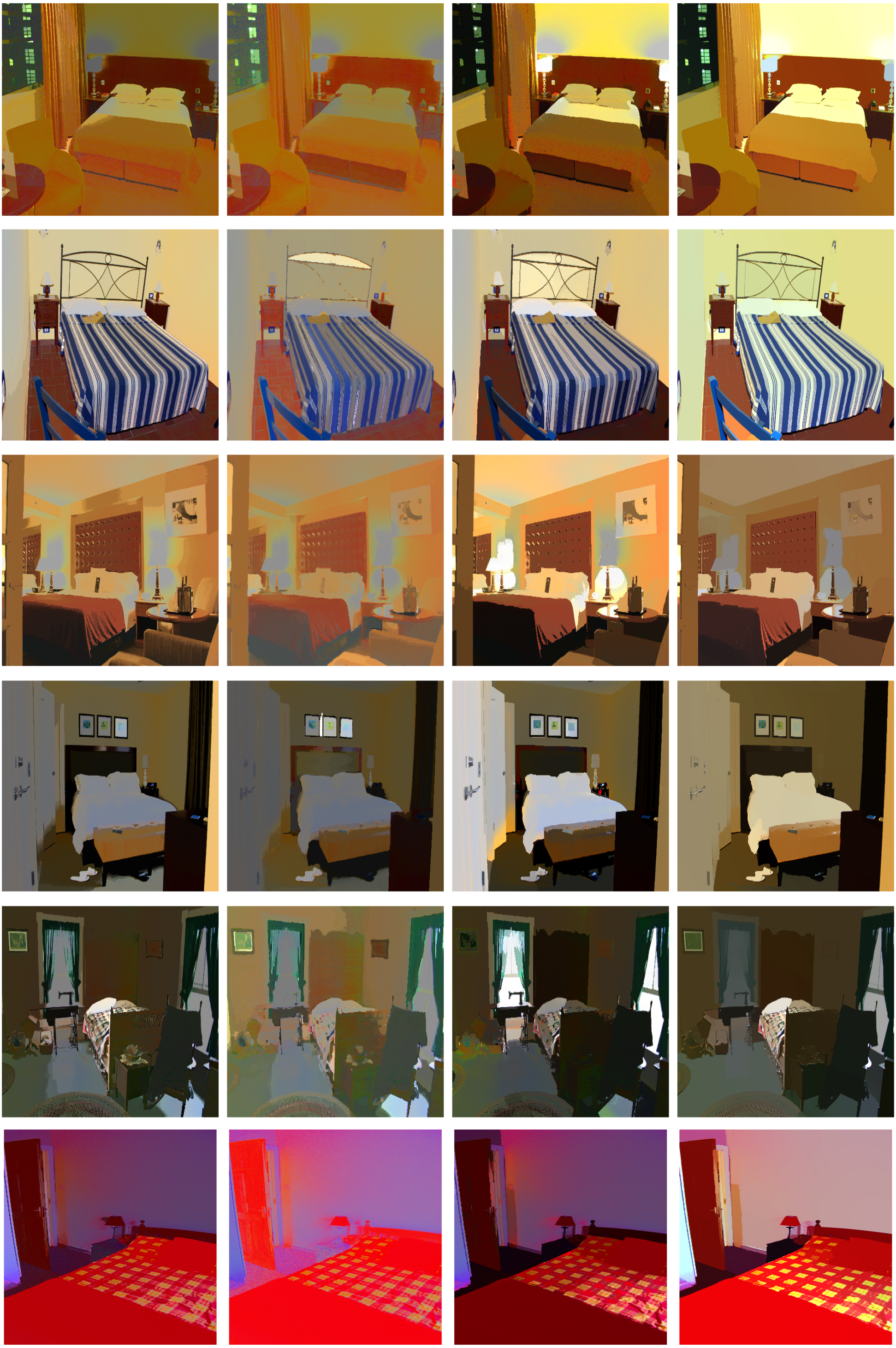

Figure 14: **Qualitative Comparison:** In each row results by [2], [4], [3] and our method respectively.

4 Failure Cases

Figure 15 shows a few failure scenarios of our proposed framework. An often observed challenging case is that of images with sharp shadow and highlight regions. Owing to the lack of depth data or some similar additional structural information, most single image IID methods struggle in this task of disambiguation of such gradients from sharp object boundaries. Yet another issue is distinguishing fine local textures in the same colour as object reflectance and lighting variation. Our method is able to handle mid-level and large textures well due to our semantic priors but in a few cases such textures get decomposed into the shading layer. Notice how in the last image the textures on the table cloth are correctly deocmposed into the reflectance layer but the textures on the wood owing to the same colour as that of the object itself, get decomposed into the shading component.

Figure 15: **Failure cases.** Note incorrect decomposition in marked regions with challenging sharp highlights, shadows and fine textures in a same colour.

5 Additional Results on Random Internet Images:

To show the generality of our method, we provide additional results on several challenging images from the Internet in Figure 16, Figure 17, Figure 18, Figure 19 and Figure 20. The images represent a variety of image types: day, night, indoors, outdoors, macro, cityscapes, paintings *etc*.

Figure 16: Additional results on random Internet images

Figure 17: Additional results on random Internet images

Figure 18: Additional results on random Internet images

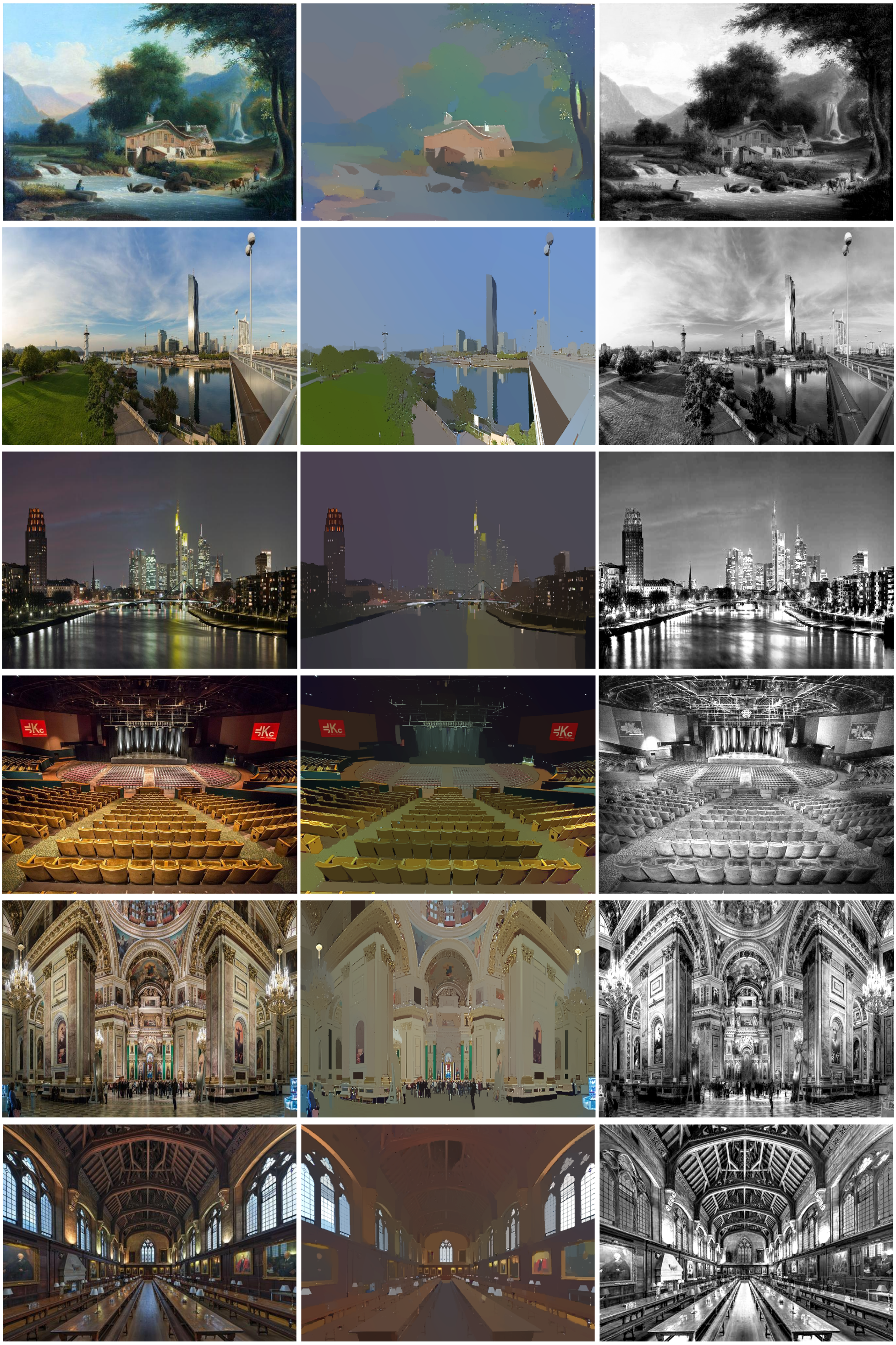

Figure 19: Additional results on random Internet images

Figure 20: Additional results on random Internet images

References

[1] P. Arbeláez, J. Pont-Tuset, J. Barron, F. Marques, and J. Malik. Multiscale combinatorial grouping. In *Computer Vision and Pattern Recognition*, 2014.

[2] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Trans. on Graphics (SIGGRAPH)*, 33(4), 2014.

[3] S. Bi, X. Han, and Y. Yu. An L_1 image transform for edge-preserving smoothing and scene-level intrinsic decomposition. *ACM Transactions on Graphics*, 34(4):78, August 2015.

[4] Tinghui Zhou, Philipp Krahenbuhl, and Alexei A. Efros. Learning data-driven reflectance priors for intrinsic image decomposition. In *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, ICCV '15, pages 3469–3477, 2015.